DOCUMENT RESUME

ED 264 266 TM 850 751

AUTHOR Tsutakawa, Robert K.
TITLE Estimation of Item Parameters and the GEM Algorithm.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Jul 82
CONTRACT N00014-81-K-0265; NR-150-464
NOTE 10p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Estimation (Mathematics); *Item Analysis; *Latent Trait Theory; *Mathematical Models; *Maximum Likelihood Statistics; Psychometrics; Statistical Studies
IDENTIFIERS *EM Algorithm; *GEM Algorithm; Parametric Analysis

ABSTRACT

The models and procedures discussed in this paper are related to those presented in Bock and Aitkin (1981), where they considered the 2-parameter probit model and approximated a normally distributed prior distribution of abilities by a finite and discrete distribution. One purpose of this paper is to clarify the nature of the general EM (GEM) solution, assuming that convergence has already taken place. For this purpose the general situation is considered and conditions are then given under which the GEM solution maximizes the likelihood function based on incomplete data. For the 2-parameter logistic model, the equations occurring at each iteration of the GEM algorithm are compared with the likelihood equations for the incomplete data. The GEM approach is shown as computationally simpler than the solution via direct methods. In practice, for latent trait applications in particular, once there is convergence, the author feels it is usually easy to test the solution by examining the likelihood function in a neighborhood of the solution and to verify whether it is at least a local maximum. This paper concludes by demonstrating that for the one parameter logistic model, convergence by the concavity of the log-likelihood function is assured. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Estimation of Item Parameters and the GEM Algorithm

ED264266

Robert K. Tsutakawa
University of Missouri

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☑ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Under the assumption that ability parameters are randomly sampled from a prior distribution with unknown parameters, various empirical Bayes type estimators of ability and item parameters were presented by Rigdon and Tsutakawa (1981). Their numerical work using the 1-parameter logistic model suggested that extensions to multiparameter models could be computationally too costly if there is interest in obtaining posterior distributions of both ability and item parameters. Here, concern will be with one of these estimators, namely an MLF procedure (a maximum likelihood procedure where the item parameters are treated as fixed parameters, as opposed to a procedure which assumes they have prior distributions), where extensions to multiparameter models appear straightforward and asymptotic methods are available for studying posterior distributions of item parameters.

The models and procedures discussed in this paper are related to those presented in Bock and Aitkin (1981), where further references may be found. Bock and Aitkin considered the 2-parameter probit model and approximated a normally distributed prior distribution of abilities by a finite and discrete distribution. They successfully used the EM algorithm by treating as unobserved data the number of examinees at each ability level and the number of these examinees who responded correctly to each item. Although they mentioned the possible use of the general EM (GEM) algorithm, they cautioned that the principle is not yet well established due to an error in a proof by Dempster, Laird, and Rubin (1977). Wu (1983) has pointed out that it was an incorrect use of the triangle inequality which invalidates Dempster, Laird, and Rubin's (1977) proof of the convergence of the EM algorithm.

One purpose of this paper is to clarify the nature of the GEM solution, assuming that convergence has already taken place. For this purpose the general situation will first be considered and conditions then given under which the GEM solution maximizes the likelihood function based on incomplete data. For the 2-parameter logistic model, the equations occurring at each iteration of the GEM algorithm will be compared with the likelihood equations for the incomplete data. It will be shown why the GEM approach is computationally simpler than the solution via direct methods. One question which remains is, under what conditions is there convergence? In practice, for latent trait applications in particular, once there is convergence, it is usually quite easy to test the solution by examining the likelihood function in a neighborhood of the solution and to verify whether it is at least a local maximum. The paper concludes by show-

TM 850 751

ing that for the 1-parameter logistic model, convergence is assured by the concavity of the log-likelihood function.

Note on the General EM Application

Dempster, Laird, and Rubin (DLR; 1977), have described the GEM algorithm for finding the maximum likelihood estimator based on the incomplete data $\underset{\sim}{y}$ when the distribution of the complete data $\underset{\sim}{x}$ does not belong to the exponential family. Following their notation, let $f(\underset{\sim}{x}|\underset{\sim}{\phi})$ be the pdf of $\underset{\sim}{x}$ depending on the r-dimensional parameter $\underset{\sim}{\phi}$ belonging to a convex set Ω in E^r and let $g(\underset{\sim}{y}|\underset{\sim}{\phi})$ be the pdf of the incomplete data, i.e.,

$$g(\underset{\sim}{y}|\underset{\sim}{\phi}) = \int_{\mathcal{X}(\underset{\sim}{y})} f(\underset{\sim}{x}|\underset{\sim}{\phi})d\underset{\sim}{x} \qquad [1]$$

where $\mathcal{X}(\underset{\sim}{y})$ denotes the set of $\underset{\sim}{x}$ mapping into the point y. Moreover, let

$$Q(\underset{\sim}{\phi}'|\underset{\sim}{\phi}) = E\{\log f(\underset{\sim}{x}|\underset{\sim}{\phi}')|\underset{\sim}{y},\underset{\sim}{\phi}\} , \qquad [2]$$

the conditional expectation of the log likelihood at $\underset{\sim}{\phi}'$ when $\underset{\sim}{\phi}$ obtains and $\underset{\sim}{y}$ is given. An instance of the GEM algorithm starts with some value $\underset{\sim}{\phi}$ and iteratively maximizes $Q(\underset{\sim}{\phi}'|\underset{\sim}{\phi})$ with respect to $\underset{\sim}{\phi}'$ while updating the value of $\underset{\sim}{\phi}$ by $M(\underset{\sim}{\phi})$, the maximizing value of $\underset{\sim}{\phi}'$ at the end of each iteration, until convergence is attained.

Suppose the algorithm converges to some value $\underset{\sim}{\phi}^*$, in the sense that the successive values of $\underset{\sim}{\phi}'$ so computed converge to $\underset{\sim}{\phi}^*$. Does $\underset{\sim}{\phi}^*$ maximize $g(\underset{\sim}{y}|\underset{\sim}{\phi})$? Since the algorithm deals with Q and not directly with g, it is not clear how the solution relates to g. This relation is clarified below.

For each $\underset{\sim}{\phi} \in \Omega$, let

$$L(\underset{\sim}{\phi}) = \log g(\underset{\sim}{y}|\underset{\sim}{\phi}) , \qquad [3]$$

$$k(\underset{\sim}{x}|\underset{\sim}{y},\underset{\sim}{\phi}) = f(\underset{\sim}{x}|\underset{\sim}{\phi})/g(\underset{\sim}{y}|\underset{\sim}{\phi}) , \qquad [4]$$

for $\underset{\sim}{x} \in \mathcal{X}(\underset{\sim}{y})$, and for each $(\underset{\sim}{\phi}', \underset{\sim}{\phi}) \in \Omega\times\Omega$, let

$$H(\underset{\sim}{\phi}'|\underset{\sim}{\phi}) = \int_{\mathcal{X}(\underset{\sim}{y})} [\log k(\underset{\sim}{x}|\underset{\sim}{y},\underset{\sim}{\phi}')]k(\underset{\sim}{x}|\underset{\sim}{y},\underset{\sim}{\phi})d\underset{\sim}{x} . \qquad [5]$$

Then, as shown by DLR, for any $(\underset{\sim}{\phi}, \underset{\sim}{\phi}^\circ) \in \Omega\times\Omega$

$$L(\underset{\sim}{\phi}) = Q(\underset{\sim}{\phi}|\underset{\sim}{\phi}^\circ) - H(\underset{\sim}{\phi}|\underset{\sim}{\phi}^\circ) . \qquad [6]$$

If the derivatives exist and $k(\underset{\sim}{x}|\underset{\sim}{y},\underset{\sim}{\phi})$ is regular in the sense of Wilks (1962), it follows from Equation 6 that

$$\left.\frac{\partial L(\underset{\sim}{\phi})}{\partial \phi_j}\right|_{\underset{\sim}{\phi}=\underset{\sim}{\phi}^\circ} = \left.\frac{\partial Q(\underset{\sim}{\phi}|\underset{\sim}{\phi}^\circ)}{\partial \phi_j}\right|_{\underset{\sim}{\phi}=\underset{\sim}{\phi}^\circ} \qquad [7]$$

since

$$\left.\frac{\partial H(\underset{\sim}{\phi}|\underset{\sim}{\phi}^{\circ})}{\partial \phi_j}\right|_{\underset{\sim}{\phi}=\underset{\sim}{\phi}^{\circ}} = 0 \qquad [8]$$

if $k(\underset{\sim}{x}|\underset{\sim}{y},\underset{\sim}{\phi})$ is regular, $j = 1, \ldots, k$.

Thus, if $L(\underset{\sim}{\phi})$ has a mode $\hat{\underset{\sim}{\phi}}$, which is also the unique solution to the likelihood equations,

$$\frac{\partial L(\underset{\sim}{\phi})}{\partial \phi_j} = 0 \ , \ j = 1, \ldots, \qquad [9]$$

then by Equation 7 the solution $\underset{\sim}{\phi}^{\circ}$ to the equation

$$\left.\frac{\partial Q(\underset{\sim}{\phi}|\underset{\sim}{\phi}^{\circ})}{\partial \phi_j}\right|_{\underset{\sim}{\phi}=\underset{\sim}{\phi}_{\bullet}} = 0 \ , \ j = 1, \ldots, r \qquad [10]$$

will also be the mode of $L(\underset{\sim}{\phi})$, i.e., $\hat{\underset{\sim}{\phi}} = \underset{\sim}{\phi}^{\circ}$.

Moreover, if the algorithm converges to $\underset{\sim}{\phi}^*$, then $M(\underset{\sim}{\phi}^*) = \underset{\sim}{\phi}^*$, and

$$Q(M(\underset{\sim}{\phi}^*)|\underset{\sim}{\phi}^*) = Q(\underset{\sim}{\phi}^*|\underset{\sim}{\phi}^*) \ ; \qquad [11]$$

therefore,

$$\left.\frac{\partial Q(\underset{\sim}{\phi}|\underset{\sim}{\phi}^*)}{\partial \phi_j}\right|_{\underset{\sim}{\phi}=\underset{\sim}{\phi}^*} = 0 \ , \ j = 1, \ldots, r. \qquad [12]$$

Thus, $\hat{\underset{\sim}{\phi}} = \underset{\sim}{\phi}^*$ under the above uniqueness condition.

Statistical Model for Item Responses

Consider the $n \times k$ matrix $\underset{\sim}{Y}$ of binary responses Y_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, k$ where $Y_{ij} = 0$ or 1 depending whether the ith examinee's response to item j is incorrect or correct. Let

$$p_{ij} = p(Y_{ij}=1|\theta_i, \underset{\sim}{\beta}_j) \qquad [13]$$

be a probability model for responses depending on θ_i, a real valued ability (or latent trait) parameter of the ith examinee, and $\underset{\sim}{\beta}_j$, a possibly vector-valued item parameter of the jth item. Given $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)^T$ and $\underset{\sim}{\beta} = (\underset{\sim}{\beta}_1, \ldots, \underset{\sim}{\beta}_k)^T$, conditional independence is assumed among the responses in $\underset{\sim}{Y}$ in the sense that the conditional joint distribution of $\underset{\sim}{Y}$, given $(\underset{\sim}{\theta},\underset{\sim}{\beta})$ is

$$p(\underset{\sim}{y}|\underset{\sim}{\theta},\underset{\sim}{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{k} p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}} \qquad [14]$$

Further assume that $\underset{\sim}{\theta}$ may be treated as a random sample from some prior distribution, $p(\theta|\underset{\sim}{\lambda})$, having parameter $\underset{\sim}{\lambda}$.

Let $\underset{\sim}{Y}_i = (Y_{i1}, \ldots, Y_{ik})$, the response to the k items by person i. Then, the unconditional probability of responses by person i is

$$p(\underset{\sim}{y}_i|\underset{\sim}{\beta},\underset{\sim}{\lambda}) = \int p(\theta_i|\underset{\sim}{\lambda}) \prod_{j=1}^{k} p(y_{ij}|\theta_i,\beta_j)d\theta_i \qquad [15]$$

and the joint probability of $\underset{\sim}{Y}$ is

$$p(\underset{\sim}{y}|\underset{\sim}{\beta},\underset{\sim}{\lambda}) = \prod_{i=1}^{n} p(\underset{\sim}{y}_i|\underset{\sim}{\beta},\underset{\sim}{\lambda}) \ . \qquad [16]$$

There are certain uniqueness problems associated with this parameterization. These problems are usually handled by placing constraints on the parameters or by reparameterization. In the interest of keeping the discussion simple, these problems will be ignored except when discussing specific models.

Estimation of Item Parameters

As formulated above, there is a sequence of n independent and identically distributed response vectors $\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n$ with likelihood function

$$\ell(\underset{\sim}{\beta},\underset{\sim}{\lambda}) = \prod_{i=1}^{n} p(\underset{\sim}{y}_i|\underset{\sim}{\beta},\underset{\sim}{\lambda}) \ , \qquad [17]$$

of the parameter $\underset{\sim}{\phi} = (\underset{\sim}{\beta},\underset{\sim}{\lambda})$. Now consider the maximum likelihood estimator $\hat{\underset{\sim}{\phi}}$, which is the value of $\underset{\sim}{\phi}$ that maximizes the likelihood function. In practice, this estimator is not trivially obtained. Even for moderate size n and k, direct maximization by Newton-Raphson type methods is likely to be too costly (cf. Bock & Lieberman, 1970). Since the joint distribution of $(\underset{\sim}{Y},\underset{\sim}{\theta})$ generally does not belong to an exponential family, a direct application of the EM algorithm is generally not feasible. Bock and Aitkin (1981) have successfully avoided this problem by approximating the prior by a known discrete distribution. The implementation of the GEM algorithm, which extends the EM algorithm to nonexponential families, was demonstrated for latent trait models by Rigdon and Tsutakawa (1981). The fact that GEM solutions are indeed maximum likelihood estimators follows under certain conditions from the general properties given in the second section of this paper. These properties will now be discussed further in terms of the 2-parameter logistic model.

Consider the 2-parameter logistic model defined by

$$p(y_{ij}|\theta_i,a_j,b_j) = \frac{e^{y_{ij}(a_j+b_j\theta_i)}}{1+e^{(a_j+b_j\theta_i)}} \qquad [18]$$

where $(\theta_1, \ldots, \theta_n)$ is a random sample from N(0, 1), $b_j > 0$ and the previous $\underset{\sim}{\beta}_j$ is now (a_j, b_j). This model is equivalent to the more familiar model

$$p(y_{ij}|\theta_i',\alpha_j,\beta_j) = \frac{e^{y_{ij}\alpha_j(\theta_i'-\beta_j)}}{1+e^{\alpha_j(\theta_i'-\beta_j)}} \qquad [19]$$

with the constraints $\Sigma_{j=1}^{k}\beta_j = 0$ and $\Pi_{j=1}^{k}\alpha_j = 1$, where $(\theta_1', \ldots, \theta_n')$ is a random sample from $N(\mu,\sigma^2)$ with both μ and σ^2 unknown. The equivalence may be seen by the relation

$$\theta_i = (\theta_i'-\mu)/\sigma \ ,$$

$$a_j = -\alpha_j(\beta_j-\mu) \ , \qquad [20]$$

and

$$b_j = \sigma\alpha_j \ .$$

The joint probability of $\underset{\sim}{y}$ is then given by

$$g(\underset{\sim}{y}|\underset{\sim}{\beta}) = \prod_{i=1}^{n}\int\prod_{j=1}^{k} p(y_{ij}|\theta_i,a_j,b_j)p(\theta_i)d\theta_i \ , \qquad [21]$$

where $p(\theta_i) = [1/(2\pi)^{1/2}]\exp(-\theta_i^2/2)$ and the posterior pdf of θ_i by

$$p(\theta_i|\underset{\sim}{y},\underset{\sim}{\beta}) = p(\theta_i|\underset{\sim}{y}_i,\underset{\sim}{\beta})$$

$$\propto p(\theta_i)\prod_{j=1}^{k} p(y_{ij}|\theta_i,a_j,b_j) \ . \qquad [22]$$

Moreover, the partial derivatives of the log likelihood are

$$\frac{\partial}{\partial a_J}\log g(\underset{\sim}{y}|\underset{\sim}{\beta}) = \sum_{i=1}^{n}\frac{\int p(\theta_i)\prod_{j=1}^{k} p(y_{ij}|\theta_i,a_j,b_j)\frac{\partial}{\partial a_J}\log p(y_{iJ}|a_J,b_J,\theta_i)d\theta_i}{g(\underset{\sim}{y}_i|\underset{\sim}{\beta})}$$

$$= \sum_{i=1}^{n} \int \left\{ y_{iJ} - \frac{1}{1 + \exp(-a_J - b_J\theta_i)} \right\} p(\theta_i | \underset{\sim}{y}_i, \underset{\sim}{\beta}) d\theta_i$$

$$= \sum_{i=1}^{n} y_{iJ} - \Sigma\, E\left\{ \frac{1}{1 + \exp(-a_J - b_J\theta_i)} \,\middle|\, \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\} \qquad [23]$$

and

$$\frac{\partial \log g(\underset{\sim}{y}|\underset{\sim}{\beta})}{\partial b_J}$$

$$= \sum_{i=1}^{n} \frac{\int p(\theta_i) \prod_{j=1}^{k} p(y_{ij}|\theta_i, a_j, b_j) \frac{\partial}{\partial b_J} \log p(y_{iJ}|\theta_i, a_J, b_J) d\theta_i}{g(\underset{\sim}{y}_i|\underset{\sim}{\beta})}$$

$$= \sum_{i=1}^{n} \int \left\{ y_{iJ}\theta_i - \frac{\theta_i}{1 + \exp(-a_J - b_J\theta_i)} \right\} p(\theta_i | \underset{\sim}{y}_i, \underset{\sim}{\beta}) d\theta_i$$

$$= \sum_{i=1}^{n} y_{iJ}\, E\{\theta_i | \underset{\sim}{y}_i, \underset{\sim}{\beta}\} - \sum_{i=1}^{n} E\left\{ \frac{\theta_i}{1 + \exp(-a_J - b_J\theta_i)} \,\middle|\, \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\} \qquad [24]$$

where $g(\underset{\sim}{y}_i|\underset{\sim}{\beta})$ is the marginal of $\underset{\sim}{y}_i$, J = 1, ..., k. Thus, the likelihood equations are

$$q_J = \sum_{i=1}^{n} E\left\{ \frac{1}{1 + \exp(-a_J - b_J\theta_i)} \,\middle|\, \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\} \qquad [25]$$

and

$$\sum_{i=1}^{n} E\{y_{iJ}\theta_i | \underset{\sim}{y}_i, \underset{\sim}{\beta}\} = \sum_{i=1}^{n} E\left\{ \frac{\theta_i}{1 + \exp(-a_J - b_J\theta_i)} \,\middle|\, \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\}, \qquad [26]$$

J = 1, ..., k, where $q_J = \Sigma_{i=1}^{n} y_{iJ}$.

On the other hand, consider the maximization of $Q(\underset{\sim}{\beta}'|\underset{\sim}{\beta})$ with respect to $\underset{\sim}{\beta}'$ in a given iteration of the GEM algorithm applied to the 2-parameter logistic model. (Note that now $\underset{\sim}{\phi} = \underset{\sim}{\beta}$ and $\underset{\sim}{\phi}' = \underset{\sim}{\beta}'$, since λ does not exist under the current parameterization.) Now

$$Q(\underset{\sim}{\beta}'|\underset{\sim}{\beta}) = E\left\{ \sum_{i=1}^{n} \log p(\theta_i) | \underset{\sim}{\beta}, \underset{\sim}{y} \right\} + E\left\{ \Sigma\Sigma_{ij} \log p(y_{ij}|\theta_i, a_j', b_j') | \underset{\sim}{y}, \underset{\sim}{\beta} \right\}. \qquad [27]$$

Setting the partial derivatives of $Q(\underset{\sim}{\beta}'|\underset{\sim}{\beta})$ with respect to a_J' and b_J' equal to zero gives the equations

$$q_J = \sum_{i=1}^{n} E\left\{ \frac{1}{1 + \exp(-a_J' - b_J'\theta_i)} | \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\} \qquad [28]$$

and

$$\sum_{i=1}^{n} E\{y_{iJ}\theta_i | \underset{\sim}{y}_i, \underset{\sim}{\beta}\} = \sum_{i=1}^{n} E\left\{ \frac{\theta_i}{1 + \exp(-a_J' - b_J'\theta_i)} | \underset{\sim}{y}_i, \underset{\sim}{\beta} \right\} \qquad [29]$$

$J = 1, \ldots, k$.

Note the similarities between the likelihood equations and the GEM equations, i.e., Equations 25 and 26 vs. Equations 28 and 29. The important difference between Equations 25 and 28 is that in Equation 28 the conditional expectation is with respect to $\underset{\sim}{y}_i$ and a fixed value of $\underset{\sim}{\beta}$ obtained from the previous iteration; whereas in Equation 25 the conditional expectation is with respect to $\underset{\sim}{y}_i$ and the variable $\underset{\sim}{\beta}$, which also occurs in the expectation, since $\underset{\sim}{\beta} = \{(a_1, b_1), \ldots, (a_k, b_k)\}$. A similar difference can be noted between Equations 26 and 29.

Note also that since $\underset{\sim}{\beta}$ is fixed in Equations 28 and 29, the value for (a_J', b_J') can be derived separately for each J, whereas with Equations 25 and 26, (a_J, b_J) must be found simultaneously for all J, since the equations for different J are related through a common $\underset{\sim}{\beta}$. It is thus seen that the simplicity of the GEM approach is due to the considerably simpler equations to be solved.

If there is convergence of the GEM algorithm to some value $\underset{\sim}{\beta}^*$, then $M(\underset{\sim}{\beta}^*) = \underset{\sim}{\beta}^*$. In this case Equations 28 and 29 are satisfied by taking $\underset{\sim}{\beta} = \underset{\sim}{\beta}^*$ and $(a_1', b_1'), \ldots, (a_k', b_k') = \underset{\sim}{\beta}^*$. Moreover, $\underset{\sim}{\beta}^*$ satisfies Equations 25 and 26. Thus, if there is a unique solution to the likelihood equations and this is the mode, then the GEM solution will be this mode.

On the Concavity of the Log-Likelihood Function

Conditions under which the GEM algorithm produces the maximum likelihood estimate $\hat{\underset{\sim}{\phi}} = (\hat{\underset{\sim}{\beta}}, \hat{\underset{\sim}{\lambda}})$ are still unknown. However, if the logarithm of the likelihood function of Equation 3 is concave and the algorithm converges to a finite value, then this value is the maximum likelihood estimate. An important example where such a property holds is the 1-parameter model

$$p_{ij} = 1/[1 + e^{-(\theta_i - \beta_j)}] , \qquad [30]$$

and when $\theta_1, \ldots, \theta_n$ are independently and identically distributed, $N(0, \sigma^2)$.

The joint distribution of the complete data $(\underset{\sim}{\theta},\underset{\sim}{Y})$ is then

$$f(\underset{\sim}{\theta},\underset{\sim}{y}|\underset{\sim}{\beta},h) = \prod_{i=1}^{n} \frac{h^{1/2}}{\sqrt{2\pi}} \exp(-h\theta_i^2/2) \prod_{j=1}^{k} \frac{e^{y_{ij}(\theta_i-\beta_j)}}{1+e^{(\theta_i-\beta_j)}}, \qquad [31]$$

where $h = \sigma^{-2}$. Now, it is well known that

$$\frac{h^{1/2}}{\sqrt{2\pi}} \exp(-h\theta_i^2/2) \qquad [32]$$

is log concave with respect to h. Moreover, it is shown by Pratt (1981, p. 105) that

$$e^{y_{ij}(\theta_i-\beta_i)} \Big/ [1 + e^{(\theta_i-\beta_j)}] \qquad [33]$$

is log concave with respect to (β_j,θ_i) for $y_{ij} = 0, 1$. Since the product of log concave functions is log concave, it follows that $f(\underset{\sim}{\theta},\underset{\sim}{y}|\underset{\sim}{\beta},h)$ is log concave with respect to $(\underset{\sim}{\theta},\underset{\sim}{\beta},h)$. According to the theorem quoted by Pratt (1981, p. 105, §3), an integral of a log concave function with respect to any of its arguments is log concave with respect to the remaining arguments. This theorem implies that

$$g(\underset{\sim}{y}|\underset{\sim}{\beta},h) = \int f(\underset{\sim}{\theta},\underset{\sim}{y}|\underset{\sim}{\beta},h)\, d\underset{\sim}{\theta} \qquad [34]$$

is log concave with respect to $(\underset{\sim}{\beta},h)$. Thus, if the GEM algorithm converges to a finite value for the 1-parameter logistic with the normal prior on ability, the resulting solution is indeed the maximum likelihood estimate.

REFERENCES

Bock, R. D., & Aitkin, M. Marginal maximum likelihood estimator of item parameters: An application of an EM algorithm. Psychometrika, 1981, 46, 443-459.

Bock, R. D., & Lieberman, M. Fitting a response model for n dichotomously scored items. Psychometrika, 1970, 35, 179-197.

Dempster, A. P., Laird, N. M., & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm (with Discussion). Journal of the Royal Statistical Society, Series B, 1977, 39, 1-38.

Pratt, J. W. Concavity of the log likelihood. Journal of the American Statistical Association, 1981, 76, 103-106.

Rigdon, S. E., & Tsutakawa, R. K. Estimation in latent trait models (Research Report 81-1; Mathematical Sciences Technical Report No. 102). University of Missouri, Department of Statistics, 1981.

Wilks, S. S. Mathematical statistics. New York: John Wiley & Sons, 1962.

Wu, C.-F. On the convergence properties of the EM algorithm. The Annals of Statistics, 1983, 11, 95-103.

ACKNOWLEDGMENTS

This study was prepared under Contract No. N00014-81-K-0265, NR 150-464, with the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research.